**AUTOMATED SCORING OF SPEECH: SELECTED REFERENCES**

**(Last updated 20 May 2023)**

Balogh, J., Bernstein, J., Cheng, J., Van Moere, A., Townshend, B., & Suzuki, M. (2012). Validation of automated scoring of oral reading. *Educational and Psychological Measurement*, *72*(3), 435-452.

Barnett, C., Green, J. R., Marzouqah, R., Stipancic, K. L., Berry, J. D., Korngut, L., ... & Yunusova, Y. (2020). Reliability and validity of speech & pause measures during passage reading in ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *21*(1-2), 42-50.

Bejar, I. I. (2011). A validity-based approach to quality control and assurance of automated scoring. *Assessment in Education: Principles, Policy & Practice*, *18*(3), 319-341.

Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *The Journal of the Acoustical Society of America*, *145*(1), 392-399.

Chen, L., & Yoon, S.-Y. (2011). Detecting structural events for assessing non-native speech. *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications,*38-45. <http://www.aclweb.org/old_anthology/W/W11/W11-14.pdf#page=50>

Chen, L., & Yoon, S.-Y. (2012). Application of structural events detected on ASR outputs for automated speaking assessment. *Proceedings of Interspeech,*767-770. <https://www.researchgate.net/publication/260593349_Application_of_Structural_Events_Detected_on_ASR_Outputs_for_Automated_Speaking_Assessment>

Chen, L., & Zechner, K. (2011). Applying rhythm features to automatically assess non-native speech. *Proceedings of Interspeech,*1861-1864.  <http://www.isca-speech.org/archive/archive_papers/interspeech_2011/i11_1861.pdf>

Chen, L., Zechner, K., Yoon, S. Y., Evanini, K., Wang, X., Loukina, A., ... & Gyawali, B. (2018). Automated scoring of nonnative speech using the speechrater sm v. 5.0 engine. *ETS Research Report Series*, *2018*(1), 1-31.

Chen, L., Tao, J., Ghaffarzadegan, S., & Qian, Y. (2018). End-to-end neural network based automated speech scoring. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6234-6238). IEEE.

Chen, M., & Zechner, K. (2011). Computing and evaluating syntactic complexity features for automated scoring of spontaneous non-native speech. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Volume 1),* 722-731. http://www.aclweb.org/old\_anthology/P/P11/P11-1073.pdf

Chen, M., & Zechner, K. (2012). Using an ontology for improved automated content scoring of spontaneous non-native speech. *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications,* 86-94. <http://www.aclweb.org/old_anthology/W/W12/W12-20.pdf#page=104>

Chen, L., Zechner, K., Yoon, S. Y., Evanini, K., Wang, X., Loukina, A., ... & Gyawali, B. (2018). Automated scoring of nonnative speech using the Speechrater sm v. 5.0 engine. *ETS Research Report Series*, *2018*(1), 1-31.

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, *51*, 14-27.

Crossley, S. A., & McNamara, D. S. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology, 17*(2), 171-192.

Downey, R., Rubin, D., Cheng, J., & Bernstein, J. (2011). Performance of automated scoring for children’s oral reading. *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 46-55). Association for Computational Linguistics.

Eni, M., Dinstein, I., Ilan, M., Menashe, I., Meiri, G., & Zigel, Y. (2020). Estimating autism severity in young children from speech signals using a deep neural network. *IEEE Access*, *8*, 139489-139500.

Evanini, K., & Wang, X. (2013). Automated speech scoring for non-native middle school students with multiple task types. *Proceedings of Interspeech,* 2435-2439. http://evanini.com/papers/evaniniWang2013toefljr.pdf

Evanini, K., Xie, S., & Zechner, K. (2013). Prompt-based content scoring for automated spoken language assessment. *Proceedings of the 8th Workshop on Innovative Use of NLP for Building Educational Applications* (pp.157-162). Association for Computational Linguistics. <http://www.aclweb.org/old_anthology/W/W13/W13-17.pdf#page=173>

Evanini, K., & Xinhao, W. (2013). Automated speech scoring for non-native middle school students with multiple task types. *Proceedings of Interspeech,*2435-2439. <http://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_2435.pdf>

Evanini, K., Hauck, M. C., & Hakuta, K. (2017). Approaches to automated scoring of speaking for K–12 English language proficiency assessments. *ETS Research Report Series*, *2017*(1), 1-11.

Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, *35*(8), 1869-1889.

Fan, J., & Yan, X. (2020). Assessing speaking proficiency: a narrative review of speaking assessment research within the argument-based validation framework. *Frontiers in Psychology*, *11*, 330. <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00330/full>

Higgins, D., Ramineni, C., & Zechner, K. (2015). The use of learner corpora for automated scoring of written and spoken responses. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 587-586)*.* Cambridge University Press.

Higgins, D., Xi, X., Zechner, K., & Williamson, D. (2010). A three-stage approach to the automated scoring of spontaneous spoken responses. *Computer Speech and Language, 25*(2), 282-306.

Higgins, D., Xi, X., Zechner, K., & Williamson, D. (2011). A three-stage approach to the automated scoring of spontaneous spoken responses. *Computer Speech & Language*, *25*(2), 282-306.

Ivanov, A., Lange, P., Ramanarayanan, V., Suendermann-Oeft, D., & Tao, J. (2016). Speed vs. accuracy: Designing an optimal ASR system for spontaneous non-native speech in a spoken dialog application. *Proceedings of the 7th International Workshop on Spoken Dialog Systems (IWSDS)*. http://www.oeft.de/su/pdf/iwsds2016.pdf

Ivanov, A., Ramanarayanan, V., Suendermann-Oeft, D., Lopez, M., Evanini, K., & Tao, J. (2015). Automated speech recognition technology for dialogue interaction with non-native interlocutors. *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue* *(SIGDIAL 2015),* 134-138. http://www.oeft.de/su/pdf/iwsds2016.pdf

Jeon, J. H., & Yoon, S.-Y. (2012). Acoustic feature-based non-scorable response detection for an automated speaking proficiency assessment. *Proceedings of Interspeech,* 1275-1278. <http://www.isca-speech.org/archive/archive_papers/interspeech_2012/i12_1275.pdf>

Khabbazbashi, N., & Galaczi, E. D. (2020). A comparison of holistic, analytic, and part marking models in speaking assessment. *Language Testing*, *37*(3), 333-360.

Klebanov, B. B., & Loukina, A. (2021). Exploiting structured error to improve automated scoring of oral reading fluency. In Roll, I., McNamara, D., Sosnovsky, S., Luckin, R., & Dimitrova, V. (Eds.), *International Conference on Artificial Intelligence in Education* (pp. 76-81). Springer, Cham.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., ... & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684-7689.

Loukina, A., Lopez, M., Evanini, K., Suendermann-Oeft, D., & Zechner, K. (2015). Expert and crowdsourced annotation of pronunciation errors for automatic scoring systems. *Proceedings of Interspeech,* 2809-2813. <http://www.isca-speech.org/archive/interspeech_2015/papers/i15_2809.pdf>

Loukina, A., Zechner, K., & Chen, L. (2014). Automatic evaluation of spoken summaries: The case of language assessment. *Proceedings of the Building Educational Applications Workshop (BEA-9)*, 68-78. https://www.researchgate.net/profile/Chee\_Wee\_Leong/publication/279515824\_Automated\_Scoring\_of\_Speaking\_Items\_in\_an\_Assessment\_for\_Teachers\_of\_English\_as\_a\_Foreign\_Language/links/5594427408ae5d8f392f61a3.pdf#page=84

Loukina, A., Zechner, K., Chen, L., & Heilman, M. (2015). Feature selection for automated speech scoring. *Proceedings of the 10th Workshop on Innovative Use of NLP for Building Educational Applications*, 12-19.  <https://aclweb.org/anthology/W/W15/W15-0602.pdf>

Loukina, A., Madnani, N., & Cahill, A. (2017). Speech-and text-driven features for automated scoring of English speaking tasks. *Proceedings of the Workshop on Speech-Centric Natural Language Processing* (pp. 67-77). Retrieved from https://aclanthology.org/W17-4609.pdf

Qian, Y., Wang, X., Evanini, K., & Suendermann-Oeft, D. (2016). Improving DNN-based automatic recognition of non-native children speech with adult speech. *Proceedings of the Workshop on Child Computer Interaction (WOCCI).* <https://www.wocci.org/2016/files/submissions/2016/wocci2016_paper_7.pdf>

Roever, C., & Ikeda, N. (2022). What scores from monologic speaking tests can (not) tell us about interactional competence. *Language Testing*, *39*(1), 7-29.

Rusz, J., Tykalová, T., Novotný, M., Zogala, D., Růžička, E., & Dušek, P. (2022). Automated speech analysis in early untreated Parkinson's disease: Relation to gender and dopaminergic transporter imaging. *European Journal of Neurology*, *29*(1), 81-90.

Shermis, M., Burstein, J., Brew, C., Higgins. D., & Zechner, K. (2015). Recent innovations in machine scoring of student and test taker written and spoken responses. In S. Lane, M. Raymond, & T. Haladyna (Eds.), *Handbook of test development* (pp. 335-354). Routledge.

Tao, J., Evanini, K., & Wang, X. (2014). The influence of automatic speech recognition accuracy on the performance of an automated speech assessment system. *2014 IEEE Spoken Language Technology Workshop (SLT)* (pp. 294-299). IEEE.

Tao, J., Chen, L., Lee, C.M. (2016). DNN Online with iVectors Acoustic Modeling and Doc2Vec Distributed Representations for Improving Automated Speech Scoring. *Proceedings of Interspeech,* 3117-3121.

Voleti, R., Liss, J. M., & Berisha, V. (2019). A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE Journal of Selected Topics in Signal Processing*, *14*(2), 282-298.

Wang, X., Evanini, K., & Zechner, K. (2013). Coherence modeling for the automated assessment of spontaneous spoken responses. *Proceedings of the 2013 Meeting of the North American Association for Computational Linguistics: Human Language Technologies (NAACL-HLT),* 814-819. <http://www.aclweb.org/anthology/N/N13/N13-1.pdf#page=852>

Wang, X., Evanini, K., Zechner, K., & Mulholland, M. (2017). Modeling discourse coherence for the automated scoring of spontaneous spoken responses. *SLaTE* (pp. 132-137). Association for Computational Linguistics. https://www.ixa.eus/sites/default/files/dokumentuak/12795/W19-27.pdf#page=165

Wang, X., Evanini, K., Qian, Y., & Mulholland, M. (2021). Automated scoring of spontaneous speech from young learners of English using transformers. *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 705-712). IEEE.

Wilson, J., Chen, D., Sandbank, M. P., & Hebert, M. (2019). Generalizability of automated scores of writing quality in Grades 3–5. *Journal of Educational Psychology*, *111*(4), 619.

Xi, X. (2008). What and how much evidence do we need? Critical considerations for using automated speech scoring systems. In C. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 102-114). Iowa State University.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2008). *Automated scoring of spontaneous speech using SpeechRater v1.0* (Research Report RR-08-62). Educational Testing Service.

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading?. *Language Testing*, *27*(3), 291-300.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2012). A comparison of two scoring methods for an automated speech scoring system. *Language Testing, 29*(3), 371-394.

Xi, X. (2013). Validity and the automated scoring of performance tests. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 452-466). Routledge.

Xi, X., Schmidgall, J., & Wang, Y. (2016). Chinese users’ perceptions of the use of automated scoring for a speaking practice test. In G. Yu & Y. Jin (Eds.), *Assessing Chinese learners of English: Language constructs, consequences and conundrums* (pp. 150-175). Palgrave McMillan.

Xie, S., Evanini, K., & Zechner, K. (2012). Exploring content features for automated speech scoring. *Proceedings of the 2012 Meeting of the North American Association for Computational Linguistics: Human Language Technologies (NAACL-HLT),* 103-111.

Xiong, W., Evanini, K., Zechner, K., & Chen, L. (2013). Automated content scoring of spoken responses containing multiple parts with factual information. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education,* (pp. 137-142). Farmington, PA: Speech and Language Technology in Education. <http://www.mkzechner.net/slate2013tjrcontent.pdf>

Xu, J., Jones, E., Laxton, V., & Galaczi, E. (2021). Assessing L2 English speaking using automated scoring technology: Examining automarker reliability. *Assessment in Education: Principles, Policy & Practice*, *28*(4), 411-436.

Yan, D., Rupp, A. A., & Foltz, P. W. (Eds.). (2020). *Handbook of automated scoring: Theory into practice*. CRC Press.

Yeung, A., Iaboni, A., Rochon, E., Lavoie, M., Santiago, C., Yancheva, M., ... & Mostafa, F. (2021). Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech-language changes in mild cognitive impairment and Alzheimer’s dementia. *Alzheimer's research & therapy*, *13*(1), 109.

Yoon, S.-Y., Chen, L, & Zechner, K. (2010). Predicting word accuracy for the automatic speech recognition of non-native speech. *Proceedings of Interspeech,* 773-776.<http://www.isca-speech.org/archive/archive_papers/interspeech_2010/i10_0773.pdf>

Yoon, S.-Y., Evanini, K. & Zechner, K. (2011). Non-scorable response detection for automated speaking proficiency assessment. *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications*, 152-160. <http://delivery.acm.org/10.1145/2050000/2043151/p152-yoon.pdf?ip=12.233.203.201&id=2043151&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=704208945&CFTOKEN=30622475&__acm__=1481739954_c71197e17f3e0379c369f84c945262ee>

Yoon, S.-Y., & Higgins, D. (2011). Non-English response detection method for automated proficiency scoring. *Proceedings of the Workshop on Innovative Use of NLP for Building Educational Applications,* 161-169. <http://www.aclweb.org/anthology/W11-1420>

Yoon, S.-Y., & Xie, S. (2014). Similarity-based non-scorable response detection for automated speech scoring. *Proceedings of the 9th Workshop on Innovative Use of NLP for Building Educational Application,* 116-123. <http://acl2014.org/acl2014/W14-18/pdf/W14-1814.pdf>

Yu, Z., Ramanarayan, V., Suendermann-Oeft, D., Wang, X., Zechner, K., Chen, L., Tao, J., & Qian, Y. (2015). Using bidirectional LSTM recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of spontaneous non-native speech. *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2015),* 338-345. http://www.oeft.de/su/pdf/asru2015.pdf

Zechner, K. & Bejar, I. I. (2006). Towards automatic scoring of non-native spontaneous speech. *Proceedings of the North American Chapter of the Association of Computational Linguistics: Human Language Technology Conference,* 216-223. http://www.aclweb.org/anthology/N/N06/N06-1.pdf#page=244

Zechner, K., Bejar, I. I., & Hemat, R. (2007). *Toward an understanding of the role of speech recognition in non-native speech assessment* (TOEFL iBT Research Report No. 02). Educational Testing Service.

Zechner, K., Chen, L., Davis, L., Evanini, K., Lee, C. M., Leong, C. W., ... & Yoon, S. Y. (2015). Automated scoring of speaking tasks in the Test of English‐for‐Teaching (TEFT™). *ETS Research Report Series*, *2015*(2), 1-17.

Zechner, K., & Evanini, K. (Eds.). (2019). *Automated speaking assessment: Using language technologies to score spontaneous speech*. Routledge.

Zechner, K., Evanini, K., Yoon, S. Y., Davis, L., Wang, X., Chen, L., ... & Leong, C. W. (2014). Automated scoring of speaking items in an assessment for teachers of English as a Foreign Language. In *Proceedings of the ninth workshop on innovative use of NLP for building educational applications* (pp. 134-142). Association for Computational Linguistics.

Zechner, K., Higgins, D., & Xi, X. (2007). SpeechRater: A construct-driven approach to scoring spontaneous non-native speech. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*. Farmington, PA: Speech and Language Technology in Education. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.384.9255&rep=rep1&type=pdf

Zechner, K., Higgins, D., Xi, X., & Williamson. D. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication, 51*(10), 883-895.

Zechner, K., Xi, X., & Chen, L. (2011). Evaluating prosodic features for automated scoring of non-native read speech. *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2011),* 461-466. <https://www.researchgate.net/profile/Lei_Chen32/publication/239765828_Evaluating_prosodic_features_for_automated_scoring_of_non-native_read_speech/links/0a85e52e44e3aac79f000000.pdf>